U DTIC FILE COPY

SECURITY CLASSIFICATION OF THIS PAGE

(4)

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION U	1b. RESTRICTIVE MARKINGS NA
	3. DISTRIBUTION/AVAILABILITY OF REPORT Distribution Unlimited
)ULE	
AD-A198 994 BER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) NA

DTIC ELECTE SEP 19 1988 S E

6a NAME OF PERFORMING ORGANIZATION California Institute of Technology 6b OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research
6c. ADDRESS (City, State, and ZIP Code) ~~Division~~ of Chemistry 147-75 California Institute of Technology Pasadena, California 91125	7b. ADDRESS (City, State, and ZIP Code) 800 North Quincy Street Arlington, Virginia 22217-5000
8a NAME OF FUNDING/SPONSORING ORGANIZATION Office of Naval Research 8b OFFICE SYMBOL (If applicable) ONR	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-86-K-0755
8c. ADDRESS (City, State, and ZIP Code) 800 N. Quincy Street Arlington, Virginia 22217-5000	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO / PROJECT NO / TASK NO / WORK UNIT ACCESSION NO

11 TITLE (Include Security Classification)
Biopolymers: Proteins and Nucleic Acids

12 PERSONAL AUTHOR(S)
John H. Richards; John N. Abelson; Peter B. Dervan; Lee H. Hood; Melvin I. Simon

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Annual	FROM 9/15/87 TO 9/14/88	September 15, 1988	16

16 SUPPLEMENTARY NOTATION

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Protein-DNA Binding/Peptide Synthesis/DNA Cleavage/ NMR of Proteins and Nucleic Acids/tRNA Specificity

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The work focuses on learning the principles that govern interactions between proteins and nucleic acids both DNA and RNA (specifically tRNA). With these principles as guides we are synthesizing peptides (of about 50 amino acids) that bind to specific regions of DNA. Various reactive functionalities are being attached to the synthetic peptides to generate reagents that cleave DNA specifically at the site to which the peptide binds.

The work also involves biophysical studies of the protein/nucleic acid complexes in order to expand our understanding of the principles of protein binding to nucleic acids. Development of improved procedures for the chemical synthesis of peptides forms another important aspect of the program.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION U
22a NAME OF RESPONSIBLE INDIVIDUAL Dr. Robert W. Newburgh & Dr. Ira Skurnick	22b TELEPHONE (Include Area Code) (202)696-4986 & 694-3145 22c. OFFICE SYMBOL ONR

DD FORM 1473, 84 MAR 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

88 9 19 009

U

Summary

The Hin recombinase system serves us as a general paradigm of the interactions between proteins and DNA. In the past year, we have been able to define, by the use of chemical protection and specific degradation techniques, those particular base pairs that interact with the recombinase protein. Further, we have determined that the protein binds to the recombination site (which has an axis of dyad symmetry) as a dimer and that half the recombination site can bind a single monomeric unit of Hin; this binding induces a bend in the DNA. Binding requires three adenine-thymine base pairs that lie in the minor groove of the DNA.

In order to generate the enormous matrix of modifications, both to the DNA and to the protein with which it interacts, genetic techniques are being developed that allow measurement over a range of eight orders of magnitude of the binding of Hin recombinase, and its structural variants, to the DNA binding site and to sites with mutant sequences.

Biophysical measurements of the binding affinity of the Hin peptide to various synthetic DNA duplexes have quantified the specificity and thermodynamics of the peptide/DNA interactions and, more importantly, have correlated these affinities with the solution structures and base sequences of the duplexes to which Hin recombinase binds *in vivo* and to closely related structural variants of these native sequences.

The insights of these studies have allowed the design and synthesis for the first time of a sequence specific DNA cleaving protein consisting entirely of naturally occurring α-amino acids.

In the area of RNA, work has focused on defining the structural principles that govern the attachment of a specific amino acid to a specific tRNA for subsequent incorporation during protein biosynthesis. This specificity determines the relationship between the sequence of bases in a structural gene and the sequence of amino acids in the protein that this gene encodes. In a particular study, a minimum of eight changes in the base composition of a tRNA specific for leucine generated a tRNA specific for serine; these changes were localized in the acceptor stem and in the dihydrouridine stem.

As peptide/protein synthesis plays a central role in this project, considerable effort has been devoted to improving the technologies for the total chemical synthesis of proteins and peptides and transferring these improved methods to other participants in the project. Of particular importance last year has been the optimization of approaches for peptide ligation that allow one to link together previously synthesized peptides of 30-40 amino acids to yield eventually proteins as large as a few hundred residues.

John Abelson

One of the projects being carried out in this laboratory has focused upon defining tRNA identity, that is those elements of a tRNA molecule which direct the correct aminoacylation of that tRNA by its cognate aminoacyl tRNA synthetase (AAS). It has been our goal to define the identity of *E. coli* serine tRNA by converting a leucine-inserting amber suppressor tRNA into one that inserts serine. We speculated that there must exist a limited number of nucleotides within a serine tRNA which determine its identity, and if superimposed upon another tRNA should be sufficient to alter the AAS recognition. We constructed a synthetic gene, $tRNA^{Leu\text{-}Ser}$, encoding an *E. coli* leucine tRNA with twelve base changes in regions presumed to be involved in AAS contact. The specificity of this mutant was determined by using the tRNA to suppress an amber mutation in the *E. coli* dihydrofolate reductase gene *fol*. Sequence analysis of the mutant protein revealed that serine and not leucine was the major amino acid being inserted by $tRNA^{Leu\text{-}Ser}$. While this

tRNA clearly had serine identity, the efficiency of suppression was quite low, 1%. Recently, we have attempted to define the minimum number of alterations required to effect the Leu to Ser conversion, speculating that some of the changes we had made were not necessary for serine identity and could have effected the efficiency of the tRNA. The contribution of each alteration was examined by reverting them to the wildtype $tRNA^{Leu}$ sequence and asking whether the resultant tRNA inserted serine. This type of analysis revealed that the minimum number of changes required to convert a leucine-inserting tRNA into a serine-inserting tRNA is eight, residing in the acceptor stem, as well as the dihydrouridine stem. This tRNA inserts exclusively serine at high efficiency (40%). We plan to continue defining the elements of serine tRNA identity, by converting other tRNAs to serine identity.

A related project has been conducted in collaboration with Jeffrey Miller and colleagues (UCLA). The aim was to expand the current collection of amber suppressor tRNA genes in *E. coli* as a means of facilitating amino acid substitution studies and protein engineering. We have constructed amber suppressor alleles corresponding to *E. coli* Phe, Cys, Pro, Gly, His, Asp, Lys, Arg. Thr, Glu, Ala, Val, Met(m) and Ile tRNAs. Seven of the new amber suppressor alleles which we have constructed, $tRNA^{Ala}_{CUA}$, $tRNA^{Cys}_{CUA}$, $tRNA^{Gly}_{CUA}$, $tRNA^{His}_{CUA}$, $tRNA^{Lys}_{CUA}$, $tRNA^{Phe}_{CUA}$, and $tRNA^{Pro}_{CUA}$ all insert the predicted amino acid. $tRNA^{Glu}_{CUA}$ inserts predominantly glutamate but is also mischarged by the glutamine AAS. Interestingly, the remainder of the suppressors, $tRNA^{Asp}_{CUA}$, $tRNA^{Arg}_{CUA}$, $tRNA^{Ile}_{CUA}$, $tRNA^{Met(m)}_{CUA}$, $tRNA^{Thr}_{CUA}$ and $tRNA^{Val}_{CUA}$, all insert lysine instead of the predicted amino acid. This result was completely unexpected, and reveals that strong identity elements or deterrents residue in the anticodon for the lysine AAS.

Publications (In Preparation)

1. Normanly, J. and Abelson, J., "A minimum of eight nucleotides are required for serine tRNA identity."

2. Normanly, J., Kleina, L. G., Abelson, J. and Miller, J. H., "Construction of *E. coli* amber suppressor tRNA. III. Determination of specificity."

3. Normanly, J. and Abelson, J., "tRNA Identity," *Ann. Rev. Biochem.*

4. Masson, J. M., Kleina, L. G., Normanly, J., Miller, J. H. and Abelson, J., "Construction of *E. coli* amber suppressor tRNA. I. Determination of suppression efficiency for 15 new suppressor genes."

5. Kleina, L. G., Normanly, J., Masson, J. M., Miller, J. H., and Abelson, J., "Construction of *E. coli* amber suppressor tRNAs. II. Improvement of suppression efficiency."

Peter Dervan

Submitted Manuscript

Mack, P., Iverson, B. L. and Dervan, P. B. Design and chemical synthesis of a sequence specific DNA-cleaving protein. *J. Am. Chem. Soc.*, submitted.

Abstract We report the design and chemical synthesis of a sequence specific DNA cleaving protein consisting wholly of naturally occurring α-amino acids. The tripeptide H-glyglyhis-OH, which is a consensus sequence for the copper binding domain of serum albumin, was attached to the amino terminus of the DNA binding domain of Hin recombinase (residues 139-190) to afford a new 55 residue protein with two structural domains each with distinct

functions, sequence specific recognition and cleavage of double helical DNA. The artificial protein was synthesized by solid-phase techniques and shown, by footprinting, to be competent to bind at 0.5 μM concentrations (pH 7.5, 25°C, 20 mM NaCl) to four Hin half sites, each 13 base pairs in length. In the presence of Cu(II) (2.5 μM), hydrogen peroxide (1 mM), and sodium ascorbate (1 mM), strong cleavage of DNA by GGH(Hin 139-190) (5 μM) occurred at one of the four sites by oxidative degradation of the deoxyribose backbone.

Completed Work - Manuscripts in Preparation

1. Sluka, J. P., Horvath, S. J., Glasgow, A. C., Simon, M. I. and Dervan, P. B. Sequence specific recognition in the minor groove of DNA by Hin protein determined by the affinity cleaving method. *Biochemistry*, in preparation.

2. Mack, D, Shin, J., Horvath, S., Simon, M. and Dervan, P. B. Orientation of the putative recognition helix for the DNA binding domain of Hin recombinase. *Biochemistry*, in preparation.

3. Sluka, J. and Dervan, P. B. Chemical synthesis of DNA binding proteins with EDTA at the amino terminus. *J. Am. Chem. Soc.*, in preparation.

4. Graham, K. and Dervan, P. B. Mapping the DNA binding domain of γδ resolvase (141-183). *Biochemistry*, in preparation.

5. Graham, K., Mack, D. and Dervan, P. B. Design of a sequence specific DNA cleaving metalloprotein. Ni GGHγδ(141-183). *Science*, in preparation.

Lee Hood (Novel Approaches to the Total Chemical Synthesis of Proteins)

Principal Investigator: Stephen B. H. Kent

Goals

The peptide synthesis portion of this grant has two purposes: to develop improved methods for the total chemical synthesis of DNA-binding proteins or active peptide fragments, and to transfer the improved methods to the other members of the collaboration.

The goal of chemical synthesis of peptides and proteins must be the production of homogeneous molecular species of defined covalent and three dimensional structure. Current methods for the chemical synthesis of the long polypeptide chains which fold to form proteins are inadequate. It is the goal of the proposed research to adapt the best aspects of existing chemistries and to develop new chemistries for the unequivocal chemical synthesis of long (100-200) polypeptide chains in pure form, suitable for use with modern spectroscopic (nmr) and diffraction techniques.

Inadequacy of Present Methods

The ultimate challenge facing synthetic peptie chemistry is the total chemical synthesis of the functional domains of proteins in pure form. A number of recent synthetic achievements, many from our own laboratory, have demonstrated that this is not beyond the realm of possibility. Automated total chemical synthesis of the 140 amino acid residue lymphokine murine IL-3 (1) and a series of analogs (2) illustrates the power and limitations of current methods. The synthetic approach to protein "engineering" definitively established the essential role of a single disulfide for the biological activity of this molecule

(2). However, it was not possible to prepare a homogeneous synthetic protein of this size using these stepwise solid phase methods (1).

Key methodology improvements which allowed this level of synthetic capability were the development of highly optimized stepwise solid phase peptide synthesis (SPPS) (4), the effective automation of this chemistry for the rapid reproducible assembly of very long protected peptide chains (5, 6), and improved procedures for the removal of protecting groups from the product (7). Empirical methods of protein folding, the use of high resolution chromatographic and electrophoretic techniques for purification and analysis, and structural characterization of the synthetic proteins by mass spectrometry and peptide mapping have also played an important role. Despite these advances, the largest proteins reproducibly chemically synthesized in pure form are only on the order ot 50 amino acid residues in length, such as insulin (8), epidermal growth factor (9), and transforming growth factor alpha (10, 11). These are the very smallest proteins. Larger, more typical proteins have so far resisted the chemical approach in terms of the synthesis of pure defined molecular species.

What are the shortcomings of existing techniqeus that have limited us to this size range? For stepwise solid phase synthesis, the main problem is lack of absoluately quantitative yields in the chain assembly. Combined with the lack of fractionation of the resin-bound intermediates, this leads to a final product contaminated with a large number of closely related molecular species, which are difficul tto separate and which interfere with the work up, folding, and characterization of the product protein. It is a measure of the power of the solid phase approach that average chain assembly yeilds of 99.4% per residue have been routinely obtianed (1) in the synthesis of a number of proteins. Nonetheless, this is not sufficient.

Progress Report

In the past year, we have continued the development of rapid, high efficiency stepwise solid phase peptide synthesis as described in last year's report. We have critically reevaluated the chemistry used for the assembly of protected peptide chains on resin supports. Swelling of peptide-resins originates from the interaction of solvents with the protected peptide and the polymer backbone. Thus, we have used a single 4 minute treatment with neat (100%) trifluoroacetic acid (TFA) to remove the Boc-group at each step of the chain assembly with 99.98% (±0.02%) efficiency. The highly swollen state of the peptide-resin results in rapid diffusion between the solution and the interior of the swollen beads, thus only three brief (35-60 second) flow washes with DMF were used in the entire synthetic cycle. Finally, very high (~0.15 M) concentrations of activated Boc-amino acids were used in the coupling step, resulting in rapid (>10 min), high efficiency formation of the peptide bond. Implementation of protocols based on these principles has resulted in a highly optoptimized, efficiency synthetic peptide chemistry, with cycle times as short as 20 minutes for the addition of each amino acid residue. A wide range of protocols (Figure 1) for manual and automated SPPS have been developed, for different scales of synthesis and for the assembly of peptide chains up to more than 100 amino acids in length. Some reprints describing our own applications of these improved synthetic protocols are enclosed.

These protocols have been transferred to the Dervan group, and have been implemented in Suzanna Horvath's laboratory and thus been made available to the rest of the members of the collaboration.

Current Work

It is our intention to develop methods for the ligation of large (30-45 residue) synthetic free (unprotected) peptides to form target long peptide chains. This approach will

maximize the theoretical advantages of fragment ligation (purification and characterization of intermediates; ready purification of the unfolded product protein), while avoiding the problems that arise with maximally protected fragments (frequent insolubility; difficult purification, characterization of the protected intermediates).

Fundamentally, we intend to take advantage of our ability to routinely synthesize, in good yield, highly puriifed free peptides 30-45 residues in length. Excellent methods exist for the purification (reverse phase HPLC; ion exchange chromagoraphy) and high resolution analytical (reverse phase HPLC; capillary zone electrophoresis) and molecular (mass spectrometry) characterization of free peptides.

To take advantage of these existing capabilitites, we need methods for the ligation of 30-45 residue peptides. Two methods will be explored: semisynthesis; and, the condensation of peptide alpha-thiocarboxylates. Preliminary data exists for both approaches in the literature, but as yet no one has combined optimized peptide synthesis with either of these methods. We propose to ndertake the following studies.

Semi-synthesis. In special instances, it is already possible to chemically engineer structural domains of proteins through semisynthetic methods (16). Until recently, almost all protein semisynthesis used fragments which were generated by cleavage of the parent molecule and subsequently modified before religation (16). Current SPPS methods are easily capable of producing large numbers of analogs of such peptides, rapidly and in high purity. For example, six analogs of the 39 amino acid peptide cytochrome C (66-104) were produced in a matter of weeks and reacted with the naturally-derived cytochrome C (1-65) homoserine lactone fragment to produce a family of mutant cytochrome C molecules with interesting redox potentials nad biochemical activities (Kent, Mascagni, Wallace, manuscript in preparation).

It is our intention to extend this approach, in collaboration with Wallace, to make entire protein molecules accessible by chemical synthesis. For example, in the case of cytochrome C, the religation of natural fragments corresponding to 1-39 and 40-104 or 40-65 has already been performed, using enzymatic ligation of an amino acid active ester to the fragment 1-38 (17). It is proposed to extend and generalize this approach as follows. The fragment 66-104 will be prepared by standard optimized SPPS. The fragment 40-65 homoserine lactone will be prepared using a novel synthetic approach which we developed, using a Boc Hse (Bzl)-Resin to directly generate the peptide Hse lactone. The fragment 1-38 will be prepared by SPPS, religated with the heme, and converted to the 1-39 active ester by enzymatic ligation.

An optimized approach of this type will open up the entire cytochrome C molecule to chemical modification at any specific sites and will allow us to introduce modified amino acids in any part of the molecule, including labelling with nmr probe nuclei or other reporter groups. This will be an important contribution to studies in this system, both for our own work on the structural origins of the peculiar properties of thsi molecule, and for others.

However, the principal limitation of the semisynthetic approach is that the religation reactions owe their efficiency to a "steric (or proximity) effect" originating in the association of fragments in special conformations. These approaches are not necessarily applicable to all systems or to any combination of fragments of a particular protein. Thus, a more general approach to the ligation of large synthetic peptides is needced.

Chemical Ligation. Perhaps the most innovative approach to the fragment ligation synthesis of large peptides has been the recent work of James Blake (18). In this approach, a peptide-resin linked is used which generates a peptide alpha-thiocarobxylate upon strong

acid cleavage from the resin. This is the only such functionality in the reacting system, and can be specifically activated for reaction with an alpha-amino group by treatment with silver ions. In a series of publications (18-20), Blake has developed and applied this approach. Particularly noteworthy is his use of differential chemical protection tactics to selectively protect/deprotect alpha- and epsilon-amino groups when both functionalities are present in a molecule. Peptides up to 92 and 104 residues in length have been unambiguously synthesized using this approach.

Two limitations of Blake's work are apparent: he has not used optimized SPPS for the high yield preparation of peptide fragments; and, more importantly, the route to the preparation of the loaded thioester resin is complex and impractical. Thus, this work has not been repeated in any other laboratory. Nonetheless, this approach has the potential to be a powerful general approach to the total chemical synthesis of protein domains.

We propose to: optimize the synthesis of amino acyl thioester-resins for the chemical synthesis of peptide alpha-thiocarobxylates by SPPS; cxplore and optimize ligation reactions using these fragments; and, to apply these methods to the total chemical synthesis of proteins for structure-function studies.

We have already begun to optimize the synthesis of the amino acid thioester resin, with encouraging results and some significant improvements over Blake's route. Once we have synthesized this substituted resin, we intend to explore the scope and limitations of the fragment ligation approach in the synthesis of large peptides which we have previously made. This will include beta-endorphin (31 residues), hepatitis B pre-S sequences (27, 44 residues), ELH (36 residues), and the 44 residue hormone GRF, which is difficult to make by stepwise SPPS.

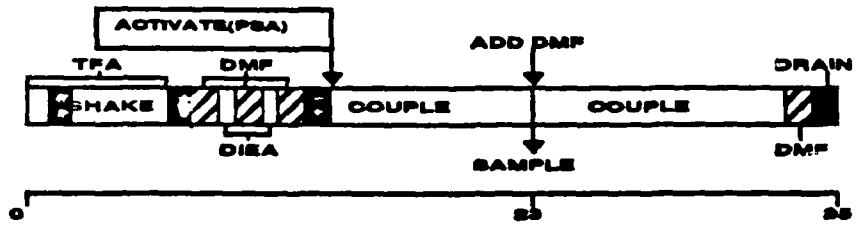
References

1. Clark-Lewis, I., Aebersold, R., Ziltener, H., Schrader, J. W., Hood, L. E. and Kent, S. B. H. (1986) *Science* **231**, 134-139.
2. Clark-Lewis, I., Lopez, A. F., Vadas, M., Schrader, J. W., Hood, L. E. and Kent, S. B. H. *Proceedings of the 5th International Lymphokine Workshop* (Pierce, C. W., Ed.), Humana Press, Clifton, New York, in press.
3. White, F. H. (1961) *J. Biol. Chem.* **263**, 1353-1360.
4. Kent, S. B. H., Clark-Lewis, I. (1985) *Synthetic Peptides in Biology and Medicine* (Alitalo, K., Partanen, P. and Vaheri, A., Eds.), Elsevier, Amsterdam, pp. 29-57.
5. Kent, S. B. H., Hood, L. E., Beilan, H., Meister, S. and Geiser, T. (1984) *Peptdies 1984: Proceedings of the 18th European Peptide Symposium*, Sweden, 1984 (Ragnarsson, Almqvist and Wiksell, Eds), Stockholm, pp. 195.
6. Kent, S. B. H., Hood, L. E., Beilan, H., Bridgham, H., Marriot, M., Meister, S. and Geiser, T. (1985) *Peptide Chemistry 1984: Proceeedings of the Japanese Peptide Symposium, 1984* (Isumiya, N, Ed.), Protein Research Foundation, Osaka, pp. 217-22.
7. Tam, J. P., Heath, W. F., Merrifield, R. B. (1983) *J. Am. Chem. Soc.* **105**, 6442-6455.
8. Seiber, P., Eisler, K., Kamber, B., Riniker, B., Rittel, W., Marki, F., De Gasporo, M. (1978) *H. S. Z. Physiol. Chem.* **359**, 113-123.
9. Akaji, K., Fujii, N., Yamima, H., Hayashi, K., Mizuta, K., Aono, M. and Moriga, M. (1985) *Chem. Pharm. Bull.* **33**, 183-201.
10. Tam, J. P., Sheikh, M. A., Solomon, D. S., Ossowski, L. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8082-8086.
11. Burgess, R. (1987) *Protein Purification: Micro to Macro* (Burgess, Ed.).
12. Kiyama, S., Fujii, N., Yajima, H., Moriga, M. and Takagi, A. (1984) *Intl. J. Pept. Prot. Res.* **23**, 174-186.
13. Fridkin, M., Patchornik, A. (1974) *Ann. Rev. Biochem.* **43**, 419-443.
14. Jakubke, H. D., Kuhl, P., Konnecke, A. (1985) *Angew. Chem. Int. Ed.* **24**, 85-93.

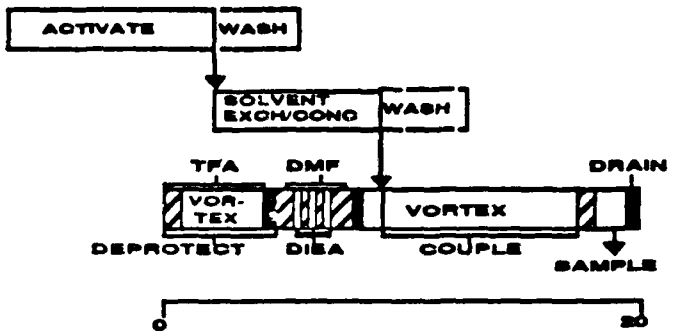
15. Nakatsuka, T., Sasaki, T. and Kaiser, E. T. (1987) *J. Am. Chem. Soc.* 109, 3808-3810.
16. Offord, R. E. (1987) *Prot. Engineering* 1, 151-157.
17. Rose, K., Herrero, C., Proudfoot, A. E. I., Offord, R. E. and Wallace, C. J. A. (1988) *Biochem. J.* 249, 83-88.
18. Blake, J. (1981) *Int. J. Pept. Prot. Res.* 17, 273-274.
19. Blake, J. Li, C. H. (1983) *Proc. Natl. Acad. Sci. USA* 80, 1556-1559.
20. Blake, J., Yamashiro, D., Ramasharma, K. and Li, C. H. (1986) *Int. J. Pept. Prot. Res.* 28, 468-476.

RAPID, HIGH EFFICIENCY SYNTHETIC PROTOCOLS

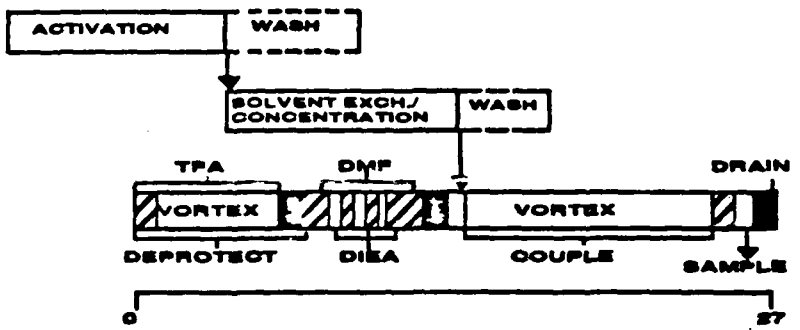
MANUAL SYNTHESIS
(0.5 mmol, 20 mL RV, 1 mmol PSA)



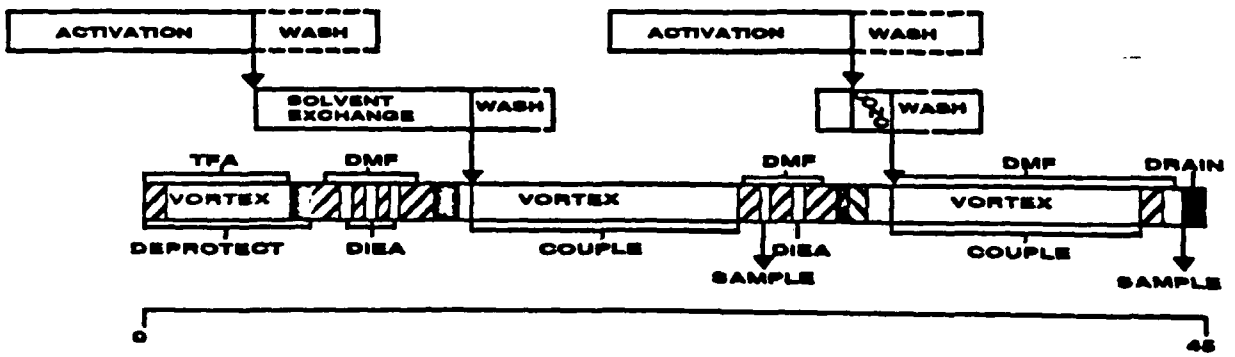
AUTOMATED SYNTHESIS
(0.1 mmol, 7 mL RV, 0.5 mmol PSA)



(0.5 mmol, 20 mL RV, 1 mmol PSA)



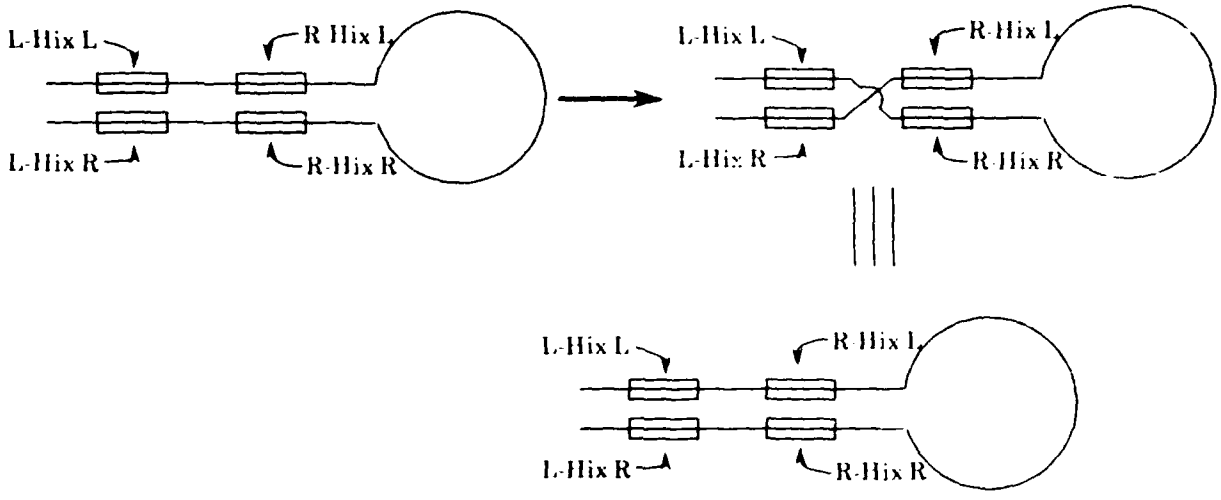
(0.3 mmol, 20 mL RV, 1 mmol PSA)



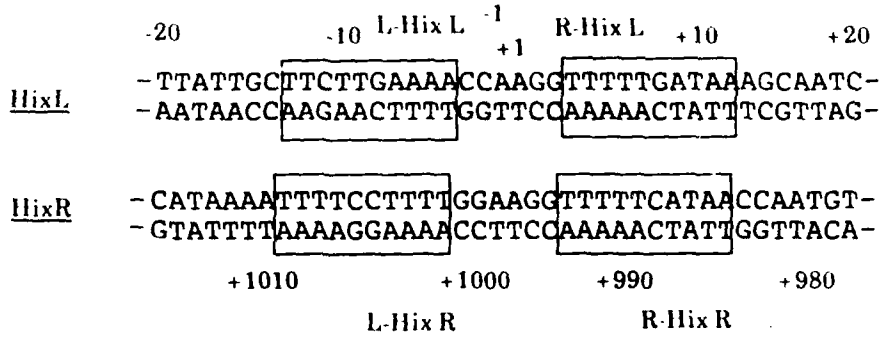
John Richards

The structural aspects of five synthetic DNA duplexes (four 14-mer duplexes and one 12-mer) duplex have been examined by 2D nmr. The affinities of these duplexes for the Hin peptide (the C-terminal 52 amino acid segment of Hin recombinase) have been determined by chromatographic techniques.

A summary of the recombination event will clarify the subseuqent discussion.



The sequences of these sites are, in detail:



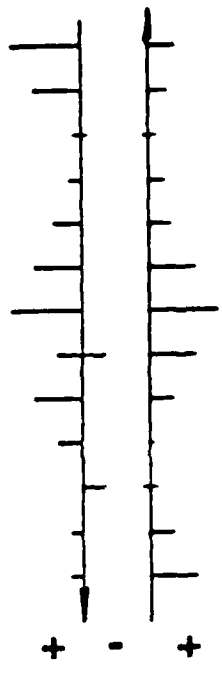
The sequences examined include:

		Binding Affinity
L-Hix L	GGTTCTTGAAACC CCAAGAACTTTGG	Strong
R-Hix L & R-HixR Common Site	GGTTTTTGATAAAA CCAAAAACTATTTC	Strong
L-Hix R	AATTTTCCTTTTGG TTAAAAGGAAAACC	Weak
Middle Sequence (between L-Hix L and R-Hix L)	CCAAGGTTTTTG GGTTCCAAAAAC	Very Weak
Self Complementary Sequence related to L-Hix L by interchange of C/G and T/A pairs	GGTTTTCGAAAACC CCAAAAGCTTTTGG	Very Weak

2D NOESY spectra were obtained with mixing time 30-500 ms. Those crosspeak intensities in the spectra with short mixing time (<100 ms) served as a basis for estimation of distances and were compared to those of fixed distances, e.g., cytosine H5-H6 at 2.54 Å (1). Because of the low S/N for these spectra, other spectra (>100 ms) were also observed; these intensities were only used for qualitative for semi-quantitative purposes.

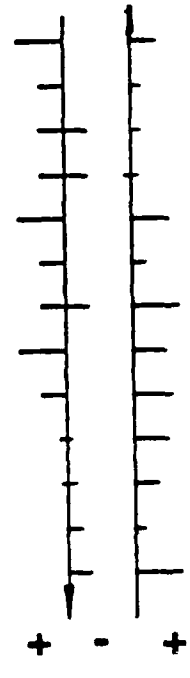
Among all the space distances, those of the interbases usually provide the most information about conformation. Among them the one that has the shortest distance is between base proton H6/8 of #N and sugar proton H2" of #(N-1), 2.1-2.2 Å in standard B-DNA (2) as in Figure 1.

Figures 2-5 show the amplitudes of these distances relative to each other. The vertical lines stand for DNA strands, the horizontal bar shows the distance differences between the measured and the calculated distances. The symbol (+) means a longer distance than that expected for B-DNA was observed; (−) signifies a shorter distance.



(L.HixL binding site)

Figure 2



(R HixL binding site)

Figure 3

Self Complementary Strands

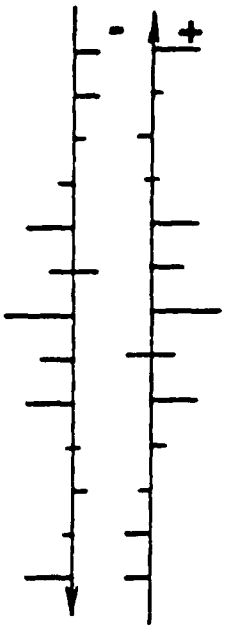


Figure 4

Central Sequence

(non-specific site)

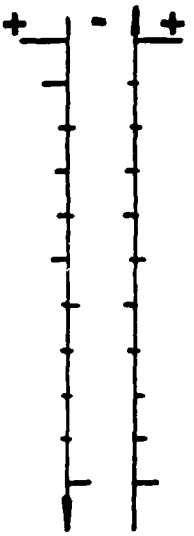


Figure 5

Conclusions

The distances between protons determined from the 2D NOESY spectra of the various duplexes and summarized graphically in Figures 2-5 show important differences in structure from classical B-DNA for the specific recognition sites of Hin recombinase. In contrast to the central sequence between L-Hix L and R, in which the interproton distances all closely approximate those for B-DNA, the L-Hix L and R-Hix L binding site sequences show significant differences, particularly between the bases in the middle of the sequence. Similar phenomena have been reported in other DNA protein complexes (3, 4). Interestingly, the self complementary sequence (Fig. 4) which has two base pairs interchanged from the sequence of L-Hix L shows similar structural distortions as do the native sequences but nevertheless does not bind the Hin recombinase peptide. This last result emphasizes that both the general backbone structure of the native DNA and putatively its ability to bend or unwind together with a very specific sequence of bases are likely requisites for strong protein/DNA binding.

References

1. Patel, D. J., Shapiro, L. and Hare, D. (1987) *Ann. Rev. Biophys. Chem.* **16**, 423-454.

2. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, p. 233.

3. Anderson, J. E., Ptashne, M. and Harrison, S. C. (1987) *Nature* **326**, 846-852.

4. McClarin, J. A., Frederick, C. A., Wang, B.-C., Greene, P., Boyer, H. W., Grable, J. and Rosenberg, J. M. (1986) *Science* **234**, 1526-1540.

Work Completed

Sun, Y., Kurutz, J. and Richards, J. H. The basis for structural and sequence specificity in protein/DNA interactions. The Hin recombinase system. *Biochemistry*, in preparation.

Melvin Simon

During the past year we have made a great deal of progress toward developing methods for engineering new polypeptides that bind and cleave DNA at specific sites. The experimental system that we are working with involves peptides derived from the Hin recombinase. Our goals have been the following:

1. To understand the nature of Hin binding to DNA. This involves knowing the specific amino acid-base pair interactions that are involved in stabilizing the binding of a subdomain of the Hin recombinase to a specific recombination site on the DNA.

2. To modify binding specificity both by being able to change specific base pairs and the cognate amino acids that interact with these base pairs in a rational way based upon our understanding of the nature of the interactions.

3. To generate peptides with new binding specificity and associate them with cleavage reagents that would break DNA with a high degree of specificity.

In order to accomplish these goals we have spent a great deal of time in the last year defining the nature of the Hin recombinase reaction and using both genetic and physical techniques to ascertain the nature of recombinase interaction with the specific site at which it acts. Dr. Anna Glasgow in our laboratory has been able to clearly define, by the use of chemical protection and specific degradation techniques, the base pairs that interact

with the recombinase protein. She has furthermore been able to show that the protein binds to the recombination site (which has an axis of dyad symmetry) as a dimer and that half the recombination site can bind a single monomeric unit of Hin. She has also shown that when the Hin recombinase dimer binds the 26 base pair recombination site it induces a "bend" in the DNA. Finally, she has developed techniques for isolating and measuring the degree of interaction between Hin recombinase and specific oligonucleotides. She has extended these measurements to fragments of the enzyme including the 52 amino acid-DNA binding domain. One of the interesting results of Dr. Glasgow's study is that she has shown that a major element required for Hin recombinase binding includes three Adenine-thymine base pairs that are in the minor groove of the DNA. Our work together with the Dervan group (see below) has defined the nature of the amino acids that are involved in some of the interactions in this minor groove binding site.

Dr. Kelly Hughes in our laboratory is in the process of developing genetic techniques to prove and modify the site and the amino acid portion of the protein that interacts with this site. Dr. Hughes has developed a number of genetic probes that allow him to measure the binding of Hin recombinase *in vivo* over a range of eight orders of magnitude. This system also allows him to modify the nature of the site and to modify the protein while measuring the relative affinities of each of the mutants to each other. This provides an enormous matrix of modifications that allows us to assess the nature of the binding interactions. In his *in vivo* work Dr. Hughes confirmed the results of chemical protection and *in vitro* studies. He showed that *in vivo* the same base pairs are involved in determining binding specificity. Dr. Hughes has been able to construct a variety of reagents that will now allow him to select changes in the polypeptide that results in the peptide acquiring the ability to interact with specific changes in the DNA binding site. His preliminary experiments indicate that these approaches will allow us to more clearly define the nature the amino acid nucleotide interactions.

In collaboration with the other members of the DARPA program group we have advanced our understanding of the binding properties of the 52 amino acid polypeptide domain derived from the Hin protein. We clarified the nature of the cutting reaction that occurs when an iron atom is attached to the terminus of this polypeptide. Using synthetically prepared polypeptides we have been able to demonstrate that three amino acids at the N terminus of the 52 amino acid polypeptide are necessary for minor groove binding. This binding represents a very important contribution to the total energy of binding of the Hin peptide. We are currently in the process of making other modifications of the peptide in the hope of changing its binding specificity.

These studies will be augmented by preliminary work which appears to be very encouraging in collaboration with Dr. Kryztopf Appelt at the Agouron Institute. We have been able to prepare small co-crystals of the 52 amino acid polypeptide and its DNA binding site. We hope that we will obtain crystals that are large enough and suitable for x-ray crystallography so that we can determine the three dimensional structure of the protein DNA complex.

In the coming year we also hope to continue our studies on the modification and mutation of the site and the binding portion of the protein. We plan to use these studies as a basis for engineering new peptides and extending our understanding of DNA binding properties of the peptide. We plan eventually to develop entirely new species of DNA binding proteins that can be used as a basis for site specific cleavage.

The following is a list of publications that have emerged from this work.

1. Bruist, M. F., Glasgow, A. C., Johnson, R. C. and Simon, M. (1987) Fis-binding to the recombinational enhancers of the Hin DNA inversion system. *Genes and Development* 1, 762-662.

2. Sluka, J. P., Horvath, S. J., Bruist, M. F., Simon, M. and Dervan, P. (1987) Synthesis of a sequence-specific DNA-cleaving peptides. *Science* **238**, 1129-1132.

3. Johnson, R. C., Glasgow, A. C. and Simon, M. (1987) Spatial relationship of the Fis binding sites for Hin recombinational enhancer activity. *Nature* **329**, 462-465.

4. Johnson, R. C. and Simon, M. (1987) Enhancers of site-specific recombination. *Trends in Genetics* **3**, 262-267.

5. Johnson, R. C., Ball, C. A., Pfeffer, D. and Simon, M. I. (1988) Isolation of the gene encoding the Hin recombinational enhancer binding protein. *Proc. Natl. Acad. Sci. USA* **85**, 3484-3488.

6. Glasgow, A. C., Hughes, K. T. and Simon, M. I. Bacterial DNA inversion systems. In *Mobile DNA* (Berg, D and Howe, M., Eds.), ASM, Washington, D. C., in press.

7. Glasgow, A. M., Bruist, M. F. and Simon, M. K. DNA-binding properties of the Hin recombinase. Submitted.

8. Hughes, K., Youderian, P. and Simon, M. I. (1988) Phase variation in *Salmonella*: Analysis of Hin recombinase and hix recombination site interaction *in vivo*. *Genes and Development* **2**, 937-948.